# Considering spatial dependence in hedonic rent price regression

**Conference Paper**

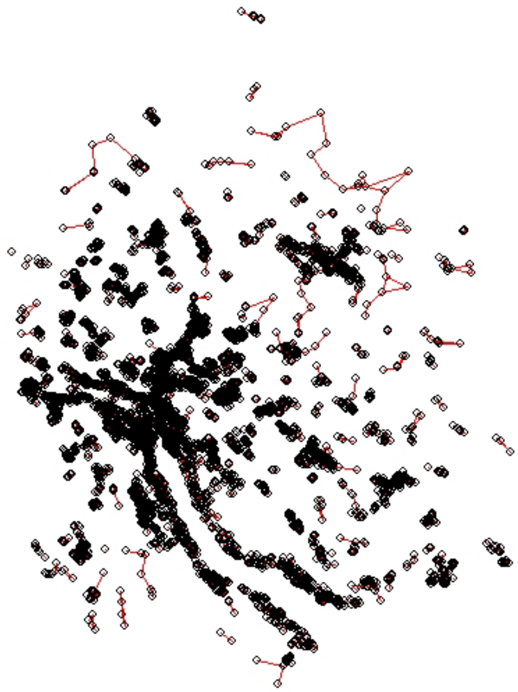
**Author(s):**
Löchl, Michael

**Publication date:**
2007

**Permanent link:**
https://doi.org/10.3929/ethz-a-005563118



**Originally published in:**
Arbeitsberichte Verkehrs- und Raumplanung 469



# Considering spatial dependence in hedonic rent price regression

**Michael Löchl, IVT/ETH**

**Conference paper STRC 2007**

STRC | 7[th] Swiss Transport Research Conference
Monte Verità / Ascona, September 12. – 14. 2007

# Improving hedonic rent price regressions by considering spatial autocorrelation

Michael Löchl
Institute for Transport
Planning and Systems (IVT)
ETH Zurich
Zurich

Phone: +41 – 44 – 633 62 58
Fax: +41 – 44 – 633 10 57
email:
loechl@ivt.baug.ethz.ch

September 2007

## Abstract

The hedonic approach has been widely used to estimate property and rent prices. However, issues of spatial effects and particularly spatial dependence on the efficiency and consistency of hedonic estimates has only recently started to receive broad attention. The analysis in this paper applies spatial regression techniques to a large Swiss dataset of bid rent prices with 9218 observations in order to improve the hedonic estimations. The reported work can be considered as preliminary results.

## Keywords

Hedonic estimation – rents – spatial dependence – spatial analysis – Zurich

# 1. Introduction

The original interest and need for real estate price data at the Institute for Transport Planning and Systems (IVT) at ETH Zurich rose from the fact that advanced land use transport models need some kind of price information in order to consider a realistic development. Particularly in cases where the necessary data is not available or the funding does not allow for buying spatially disaggregated datasets with price information, the modellers might be urged to gather and/or generate the information by themselves. This includes not only acquiring data records with price information but potentially also further spatial analysis and modelling. Particularly for public agencies and research institutes it might be possible to receive the required additional spatial data at an affordable price.

In this paper, an estimation of a hedonic rent price model for Canton Zurich is reported and further improved by considering spatial dependency. The hedonic approach has been developed by Rosen (1974), Lancaster (1966, 1971), and others, and it has been employed extensively in the study of housing values and rents. The hedonic approach looks at the price or rent as being determined by the attributes and characteristics of the dwelling unit and the surrounding neighbourhood. The hedonic regression methodology recognizes that housing is a composite product. While the attributes are not sold separately, regressing these attributes on the sales price of the composite product yields the marginal contribution of each attribute to the sales price.

Empirical work has produced substantial lists of attributes and characteristics of the welling unit to be considered (Sirmans *et al.*, 2005) and they can be roughly divided into structural and locational attributes. Structural attributes describe the physical structure of a residential unit and comprised characteristics such as size, number of rooms, condition and equipment of the residential unit, age of building etc.. Locational attributes include the surrounding area and locational externalities. Those are neighbourhood characteristics such as densities, distances to infrastructure, accessibility and others.

Many studies haven been published about solving specific econometric issues in the application of hedonic real estate price models. These issues have been classified into four categories of concern (Kim *et al.*, 2003, 24): functional form, identification, statistical efficiency and benefit estimation.

Although, there has been some earlier work (i.e. Can, 1990), the importance of spatial effects and particularly spatial dependence on the efficiency and consistency of hedonic real estate estimates has only recently started to receive broad attention (Kim *et al.*, 2003, Dubin *et al.*, 1999, Pace *et al.*, 1998, Basu and Thibodeau, 1998). The literature reveals that the neglect of spatial considerations in econometric models may lead to biased coefficients. It may even lead to errors in the interpretation of tests for heteroskedasticity (Anselin, 1988, 121). The analysis in this paper applies the hedonic approach to a Swiss dataset of bid rent prices and considers spatial regression techniques.

Spatial dependence, or more specifically spatial autocorrelation, is the correlation among values of a single variable strictly attributable to the proximity of those values in geographic space, introducing a deviation from the independent observations assumption of classical regressions (Griffith, 2003, 3).

As Griffith (2003, 5) points out, social science variables tend to be moderately positively spatially autocorrelated because of the way phenomena are geographically organized. This is particularly true for real estate markets, where dwelling units in the same neighbourhood capitalize shared location amenities, such as favourable neighbourhood characteristics (education, income and nationality of inhabitants, density of the neighbourhood), accessibility and proximity externalities (distance to commercial properties, noise pollution etc.).

## 2. Data

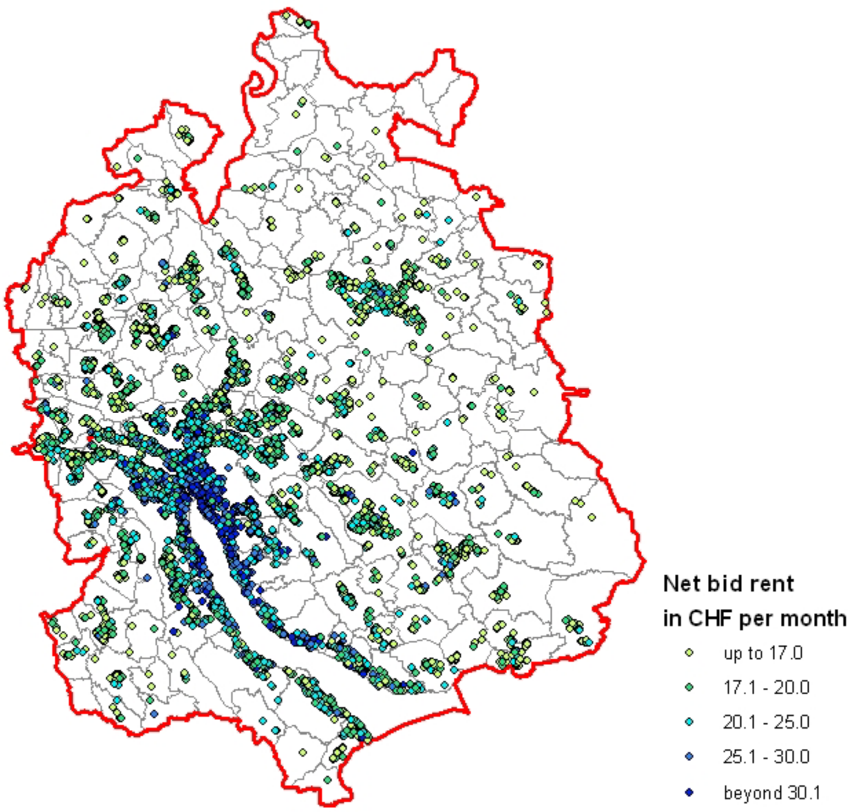
In this study, the focus has been given to residential rents. The underlying data was parsed from the comparis webpage in December 2004 as well as April and October 2005. Afterwards, dublicates were deleted. The complete dataset comprehends adds from several online databases such as Homegate, ImmoClick, ImmoScout24 and SwissImmo. The addresses were available for all records and a geocoding gave the opportunity to add more spatial information by a Geographic information system (GIS), for which *ArcGIS* 9.2 and *Python scripting* has been extensively employed.

Moreover, the interactive data-analysis environment *R* has been used for the data analysis besides some initial testing in *SPSS* Version 14. Basic information on the *R* programming environment itself may be found in the initial source (Ihaka and Gentleman, 1996) and the project site[1]. For a technical introduction into the R package *spdep* and available functions for spatial data analysis, see Bivand (2002).

Overall, the dataset comprehends rent prices and additional information of 9218 dwelling units in Canton Zurich. However, as the dataset include bid rent prices, it does not necessarily reflect paid market prices. Furthermore, the sample might be slightly biased because some vacant dwelling units do not make their way to online platforms. But an earlier comparison (Löchl, 2006) showed that the differences in the structural variables to the Federal Building and Apartment Register (GWR) of the Federal Statistical Office are not very large. Finally, a small bias might arise from the fact that the data as been collected during 10 months, although it is treated here as a cross-sectional dataset. The price information is the net bid rent, as displayed in Figure 1.

[1] http://www.r-project.org

Figure 1 Observations ($N = 9218$)




Source: Administrative boundaries from Vector25 © 2006 swisstopo (DV033492.2)

Besides the price information and the dwelling unit size in square meter, the amount of rooms is available. A dummy variable of wether it is a single family house could not be used in the model estimations since this information is missing for some 1500 observations. Moreover, the dataset includes dummy variables with information about available facilities such as balcony, chimney, and garden terrace in the dwelling unit as well as a lift in the building,

The address for every dwelling unit in the dataset has been geocoded at building level and matched with a wide set of spatial variables. The generation of some of those variables included significant further work, others were just matched with available layers by *ArcGIS*. Two examples of generated variables dealing with the terrain topography are described in more detail below.

## 2.1 Solar exposure

As an update and improvement to an earlier idea for considering solar exposure in hedonic real estate modelling (Löchl *et al.*, 2007), several solar exposure indices have been calculated based on the 25 meter digital elevation model (DEM25) by swisstopo and a new dataset including detailed altitude and elevation information of the sun over the year from the Astronomischen Institut of the University of Bern. Therefore, only the terrain is considered for the potential solar exposure while ignoring buildings as well as vegetation and a yearly average has been calculated (see Figure 1 for examples). For those calculations, the terrains of the bordering Swiss regions have been considered as well.

Figure 2 Morning (left) and evening (right) solar exposure index

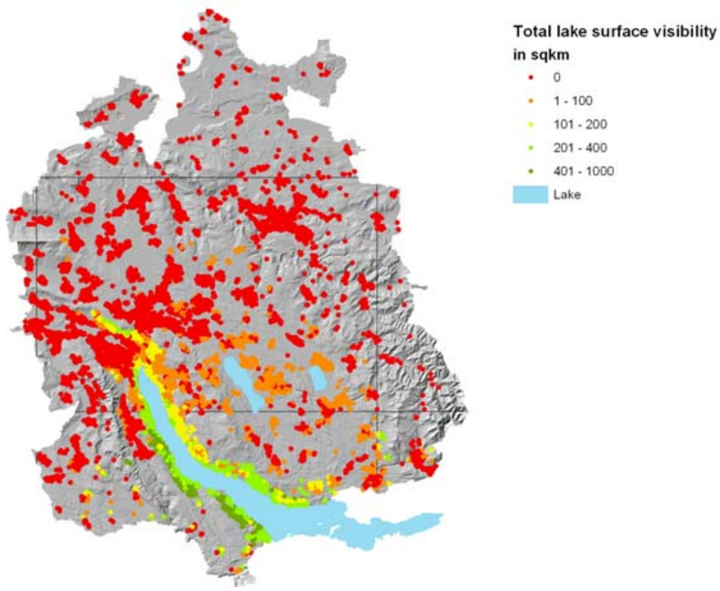

Source: DEM25 © 2006 swisstopo (DV033492.2)

## 2.2 Visibility

Particularly mountainous areas provide locations with a beautiful wide view on creeks, lakes and mountains. Since most parts of the Canton Zurich are rather hilly areas, the variance among properties in the region varies quite dramatically. Moreover, the significance of the view on the real estate price in Switzerland and beyond has been shown in other studies before (Baranzini and Schaerer, 2007; Salvi *et al.*, 2004; Rieder 2005; for an international overview, see Bourassa *et al.,* 2003).

The surface visibility has been calculated for both general cleared view and view of lake surfaces larger than 1 sqkm. For those variables, the viewshed from each property has been calculated based on the aforementioned DEM25 with GIS functionality. It considers an horizontal angle of 360 degrees, measured with an offset of 4 meters while natural cover and buildings are neglected. Therefore, the measure can only be considered as the potential view.

Figure 3 Lake view from dwelling unit (only lakes >1sqkm considered)




Source: DEM25 © 2006 swisstopo (DV033492.2)

# 3. Traditional hedonic model

In a first step, a hedonic regression model by means of ordinary least squares (OLS) has been estimated. The literature has gravitated towards the use of flexible functional form such as the Box-Cox transformation, but such specifications are not readily implemented in the presence of spatial dependence (Kim *et al.*, 2003, 31) and are therefore disregarded here. The OLS model has been estimated with a semi-log specification.

A wide set of variables have been tested. However, only a selection of them proved to be significant while controlling for multicollinearity by considering the combination with the lowest Variance Inflation Factor (Maddala, 2001, 272). Consequently, various variables had to be disregarded, for example the information about dwelling unit size and potential public transport accessibility. A descriptive summary of the variables included in the OLS model are given in Table 1. Note that transformations of variables in the model are disregarded in the table.

Table 1 Descriptive statistics

| Variable | Description | Type[1] | Min | Max | Mean | S.D. |
|---|---|---|---|---|---|---|
| Dependent | | | | | | |
| SQMRENT | Monthly rent price per square meter in CHF | C | 7.30 | 58.62 | 20.73 | 5.53 |
| Independent | | | | | | |
| ROOMS | Number of rooms | C | 1.00 | 10.00 | 3.67 | 1.24 |
| LIFT | Building has a lift | D | 0.00 | 1.00 | 0.23 | 0.42 |
| CHIMNEY | Dwelling unit has a chimney | D | 0.00 | 1.00 | 0.03 | 0.18 |
| BALCONY | Dwelling unit has at least one balcony | D | 0.00 | 1.00 | 0.45 | 0.50 |
| GTERRACE | Dwelling unit has a garden terrace | D | 0.00 | 1.00 | 0.01 | 0.12 |
| TTIME_ZH | Travel time to Buerkliplatz (ZH) by car in minutes | C | 8.00 | 58.40 | 29.88 | 9.18 |
| RSTATION | Euclidean distance to next rail station in meters | C | 0.01 | 5.73 | 0.91 | 0.66 |
| RAIL | Rail line within 50 m | D | 0.00 | 1.00 | 0.04 | 0.20 |
| AUTOBAHN | Autobahn within 100 m | D | 0.00 | 1.00 | 0.02 | 0.14 |
| AUTOEXIT | Next autobahn exit within 2 km | D | 0.00 | 1.00 | 0.59 | 0.49 |
| AIRNOISE | Daily average of air noise is above 52dB | D | 0.00 | 1.00 | 0.07 | 0.25 |
| SOLAR_EVE | Evening solar exposure index | C | 0.00 | 6.16 | 2.38 | 0.81 |
| VIEW_ALL | Total clear visibility of terrain surface in sqkm | C | 0.91 | 784.21 | 88.42 | 55.96 |
| VIEW_LAKE | Visibility of lake surface (>1 sqkm) in sqkm | C | 0.00 | 88.88 | 4.46 | 10.58 |
| POP_DENS | Number of inhabitants in hectare | C | 0.00 | 496.00 | 90.84 | 61.38 |
| CATERING | Number of jobs in catering industry per hectare within 1 km | C | 0.00 | 14.73 | 0.61 | 1.57 |
| FOREIGNER | Percentage of foreigners[2] in hectare | C | 0.00 | 50.00 | 4.98 | 3.74 |
| GROCERY | Grocery store (≥ 400 sqm) within 500 meters | D | 0.00 | 1.00 | 0.46 | 0.50 |
| CONSTRUCT | Percentage of buildings built before 1971in municipality | C | 0.00 | 79.61 | 59.09 | 16.34 |
| INCOME | Income per capita in municipality in 1000 CHF in 2003 | C | 23.99 | 72.18 | 35.71 | 7.21 |

[1] C = continues; D = dummy

[2] Inhabitant with nationalities outside of North-Western Europe, North America and Australia

As can be observed in Table 3, the OLS model with the normal logarithmic transformation of the monthly rent price per square meter as a dependent variable achieves a reasonable fit ($R^2$ = 0.511 for $N$ = 9218), considering the relatively few included structural variables and the fact that the data is cross-sectional and not based on longitudinal data. All estimated coefficients have the expected sign as described below.

The large amount of significant spatial variables in the OLS model highlights the importance of the location of a property. Even without any structural variable in the estimation, a $R^2$ of 0.449 can be reached.

The number of rooms has been included with a radical transformation, as suggested by the literature (Scognamiglio, 2002, 78). All other available structural (dummy) variables proved to be significant with a positive coefficient, namely the availability of balcony, chimney, garden terrace and lift.

By far the most important variable is the normal logarithmic travel time to the inner city (Buerkliplatz) of Zurich, which highlights the importance of the city for the region. Because of multicollinearity issues, no other regional accessibility measure could be included in the estimation. Instead, the variable of the normal logarithmic distance to the next rail station as the local node and backbone of the public transport system in the canton proved to be useful. The noise disturbance from transport plays a prominent role in the final OLS model, as the availability of a rail line (within 50 meters) or an autobahn (within 100 meters) in close proximity reduces the price. The same can be observed for air noise, at least as long as it is below a certain level (>52dB). Surprisingly, the availability of an autobahn exit in close proximity reduces the price, since usually one would expect that it improves the access and therefore increases the rent price. However, good access may attract undesirable (car dependent) land uses, as well as a relatively high density of traffic to and from the autobahn. Some variables which are taking into account terrain topography in some form are included as well. From the calculated solar exposure indices as described in section 2.1, the evening solar radiation index turned out to be significant with a positive coefficient. Positively valued is the logarithmic transformation of the view of both the total terrain surface and the lake surface (only considering lakes larger than 1 sqkm). High neighbourhood density, measured by the number of inhabitant in the hectare (as in Swiss Census 2000) has a negative effect on the price while the density of restaurants and hotels (measured by the number of referring jobs per hectare within 1 km) has a positive effect. The percentage of foreign inhabitants from countries outside North-Western Europe, North America and Australia in a hectare (as in Swiss Census 2000) is valued negatively. The information of the presence of at least one grocery store (≥ 400 sqm) within 500 meters is taken from the “Betriebszählung 2005” by the Swiss Statistical Office (BFS) by considering the middle point of the referring hectare and the measure has therefore an accuracy of +- 50 meters. This dummy has a positive coefficient. Finally, two variables at municipality level are included, both with a positive coefficient. One is the percentage of buildings built before 1971, the other one is the normal logarithmic income per capita.

After the OLS estimation, a set of diagnostics for spatial autocorrelation have been performed, which are explained in the next section. However, they clearly indicated the need to consider spatial autocorrelation in the dataset. This is no surprise, considering that most variables in the estimation have actually a spatial relation. Anselin (1988, 121) points out that in case of spatial dependence, tests for heteroscedasticity are no longer valid. Consequently, they are disregarded at this point of the analysis.

# 4. Spatial hedonic models

In general, there are two ways to incorporate spatial effects into a regression model. While spatial lag models pertain to spatial correlation in the dependent variable, the latter refers to the error term. If spatial autocorrelation is not considered when it is in fact present, it would have different consequences, depending on whether the correct model is a spatial lag or a spatial error specification. Anselin (2006b, 13) describes it as equivalent to an omitted variable when a spatially lagged dependent variable is ignored, which would yield OLS estimates for the model coefficients that are biased and inconsistent. The case of ignoring spatially correlated errors is mostly a problem of efficiency, in the sense that the OLS coefficient standard error estimates are biased, but the coefficient estimates themselves remain unbiased. However, to the extent that the spatially correlated errors mask an omitted variable, the consequences of ignoring this may be serious. Both models are described in more detail in the following.

## 4.1 Spatial lag model

The spatial lag model is also called spatial simultaneous autoregressive lag model (SAR). A spatial-lag hedonic rent price model can be written as follows:

$$\mathbf{P} = \rho\mathbf{W} + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \varepsilon, \qquad (1)$$

where $\mathbf{P}$ is the vector of rent prices, $\rho$ is a spatial autocorrelation parameter, $\mathbf{W}$ is a $n \times n$ spatial weight matrix (where $n$ is the number of observations), $\mathbf{X}_1$ is a matrix with observations on structural characteristics, $\mathbf{X}_2$ is a matrix with observations on location characteristics, with $\varepsilon$ assumed to be a vector of independent and identically distributed error terms. Typically, the definition of neighbours used in the weights matrix is based on a notion of distance decay or contiguity.

## 4.2 Spatial error model

In case when spatial dependence is present in the error term, a spatial autoregressive specification for this dependence is usually assumed. This is called spatial error model (SEM) und can be formulated as follows:

$$\mathbf{P} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + u,$$

$$u = \lambda\mathbf{W}\,u + \varepsilon, \qquad (2)$$

where $\lambda$ is the spatial autoregressive coefficient, **W** is the spatial weight matrix, and $u$ is assumed to be a vector of identically distributed errors. This model is a special case of a regression specification with a non-spherical error variance-covariance matrix. Therefore, W now pertains to shocks in the unobserved variables (the errors $u$) but not to the explanatory variables of the model (**X**). Consequently, the price at any location is a function of the local characteristics but also of the omitted variables at neighbouring locations.

## 4.3 Contiguity matrices

In incorporate spatial dependence, it is necessary to produce a weight matrix. There are several approaches to define contiguity (i.e. Anselin, 2002, 258). Because of a relatively high heterogeneity of spatial distribution of the data points, *a k-nearest neighbours* approach (by euclidean distance) has been chosen. Therefore, each unit has the same number of neighbours, but the weights matrix becomes asymmetric. However, this can be easily corrected by the *make.sym.nb* function in *R*. Usually, the diagonal elements of the weights matrix are set to zero and row elements are standardized such that they sum to one. With this row-standardized spatial weights matrix, it amounts to including the average of the neighbours as an additional variable into the regression specification (Anselin, 2006b).

In order to detect spatial autocorrelation in the OLS model, Moran's I and the Lagrange Mulitplier Test have been used by considering only the next neighbour ($k$ = 1). While the Moran's I statistic has power in detecting misspecifications (even beyond the problem of spatial autocorrelation), it does not provide sufficient information of which alternative specification should be used (Anselin, 2005, 197). For this task, the Lagrange Multiplier test statistics are more helpful. While a significant Moran's I value of 0.674 for the dependent variable Ln(SQMRENT) clearly indicated spatial autocorrelation, the Lagrange Multiplier Test statistics suggested to work further with rather a spatial error model as the Robust LM-Error statistic is highly significant (with $p < 0.001$) while the Robust LM-Lag statistic was only significant at a lower level (with $p = 0.02$). However, both the SEM and SAR model has been considered in the further work.

Often, the 8 next neighbours are used for the *k-nearest neighbours* approach, but this choice does not always follow a straightforward approach. The literature is suggesting that different weight matrices may alter results significantly and to test the results for different specifications. Here, model fit tests with matrices of 1 to 16 next neighbours have been performed for both the SEM and the SAR model in order to define the most appropriate $k$ for the further work. The results are given in Table 2. The traditional $R^2$ measure of fit, based on the decomposition of total sum of squares into explained and residual sum of squares, is not applicable here (Anselin, 1992, 190). Instead, the so-called pseudo-$R^2$ has been used, as recommended by Anselin:

$$R^2 = \frac{\sigma^2_{pred}}{\sigma^2_{obs}} \tag{3}$$

where $\sigma^2_{pred}$ is the variance of the predicted response variable and $\sigma^2_{obs}$ is the variance of the observed response variable. Moreover, the Akaike Information Criterion (AIC, Bozdogan, 1987) and the log likelihood have been used as measures for goodness-of-fit. The model with the highest log likelihood, or with the lowest AIC is best (Anselin, 1992, 190). Finally, the Moran's I for the dependent variable Ln(SQMRENT) has been added as well.

Table 2 Measures of quality of fit for spatial regressions for different weighting matrices

| *k* | Morans's I of Ln(SQMRENT) | SEM Pseudo $R^2$ | SEM Log Lik. | SEM AIC | SAR Pseudo $R^2$ | SAR Log Lik. | SAR AIC |
|---|---|---|---|---|---|---|---|
| 1 | 0.674 | 0.535 | 3723 | -7402 | 0.538 | 3707 | -7367 |
| 2 | 0.632 | 0.536 | 3742 | -7439 | 0.538 | 3725 | -7405 |
| 3 | 0.610 | 0.536 | 3740 | -7435 | 0.539 | 3720 | -7393 |
| 4 | 0.595 | 0.535 | 3729 | -7413 | 0.539 | 3713 | -7381 |
| 5 | 0.582 | 0.535 | 3725 | -7405 | 0.539 | 3710 | -7375 |
| 6 | 0.571 | 0.533 | 3708 | -7370 | 0.538 | 3697 | -7347 |
| 7 | 0.561 | 0.532 | 3695 | -7343 | 0.538 | 3688 | -7331 |
| 8 | 0.553 | 0.531 | 3690 | -7334 | 0.538 | 3688 | -7331 |
| 9 | 0.546 | 0.530 | 3671 | -7297 | 0.537 | 3669 | -7292 |
| 10 | 0.540 | 0.529 | 3661 | -7277 | 0.537 | 3662 | -7279 |
| 11 | 0.535 | 0.527 | 3641 | -7237 | 0.536 | 3646 | -7246 |
| 12 | 0.530 | 0.526 | 3624 | -7202 | 0.535 | 3629 | -7213 |
| 13 | 0.525 | 0.525 | 3607 | -7168 | 0.534 | 3614 | -7183 |
| 14 | 0.522 | 0.524 | 3605 | -7163 | 0.534 | 3614 | -7183 |
| 15 | 0.518 | 0.524 | 3597 | -7147 | 0.533 | 3607 | -7169 |
| 16 | 0.515 | 0.523 | 3588 | -7130 | 0.533 | 3599 | -7153 |

In general, the results suggest that with an increasing number of neighbours the fit decreases for both the SEM and the SAR model. This is rather unusual because often the best fits are achieved with more neighbours (i.e. Hackney *et al.*, 2006), a reason why the next 8 neighbours are mostly considered. But because of the mentioned heterogeneity of the spatial distribution of observations the result makes sense, since the next neighbour might be too far away. For both the SEM and the SAR model, highest log likelihood and lowest AIC values are achieved with the matrix of 2 next neighbours, while at least for the SAR model the matrix with 3 next neighbours has the highest $R^2$. Nevertheless, the next 2 neighbours matrix has been used for the spatial models.

The results are given in Table 3 and they show only few differences in the coefficients between the OLS model and the spatial models with the exception of a couple of variables in the SAR model. There, the constant is considerably lower than in the OLS model. Conversely, the coefficients for both the normal logarithmic travel time to Zurich (Ln(TTIME_ZH) and income (Ln(INCOME) are considerably higher, however the signs stay the same.

The spatial models indicate a higher $R^2$ of 0.536 for the SEM and 0.538 for the SAR model than in the OLS model. This relatively small difference to the non-spatial model can be explained with the broad set of spatial variables, which are already included in the OLS model. Concerning heteroskedasticity, a visual control of the residuals shows a relative homogeneity of their variance. However, the Breusch-Pagan test is significant for all models, including the OLS model. Although beyond the scope of this paper, different specifications should be tested in further research. The literature suggests a generalized least squares or weighted least squares approach in those cases (i.e. Fletcher *et al.*, 2000).

Table 3 Estimated model parameters

| Variable | OLS coeff. | t statistic | sign. | VIF | SEM coeff. | sign. | SAR coeff. | sign. |
|---|---|---|---|---|---|---|---|---|
| (Constant) | 1.716 | 11.050 | *** | | 1.740 | *** | 1.298 | *** |
| (ROOMS)^0.5 | -0.181 | -32.688 | *** | 1.194 | -0.180 | *** | -0.180 | *** |
| LIFT | 0.025 | 5.534 | *** | 1.115 | 0.018 | *** | 0.020 | *** |
| CHIMNEY | 0.104 | 10.301 | *** | 1.074 | 0.091 | *** | 0.094 | *** |
| BALCONY | 0.021 | 5.619 | *** | 1.151 | 0.022 | *** | 0.019 | *** |
| GTERRACE | 0.079 | 5.174 | *** | 1.005 | 0.072 | *** | 0.073 | *** |
| Ln(TTIME_ZH) | -0.263 | -33.102 | *** | 2.363 | -0.264 | *** | -0.180 | *** |
| Ln(RSTATION) | -0.013 | -4.930 | *** | 1.255 | -0.014 | *** | -0.010 | *** |
| RAIL | -0.030 | -3.253 | *** | 1.051 | -0.030 | ** | -0.026 | ** |
| AUTOBAHN | -0.048 | -3.677 | *** | 1.034 | -0.047 | ** | -0.038 | ** |
| AUTOEXIT | -0.035 | -8.630 | *** | 1.285 | -0.035 | *** | -0.024 | *** |
| AIRNOISE | -0.039 | -5.274 | *** | 1.088 | -0.039 | *** | 0.025 | *** |
| SOLAR_EVE | 0.026 | 11.646 | *** | 1.069 | 0.024 | *** | 0.019 | *** |
| Ln(VIEW_ALL) | 0.005 | 2.168 | ** | 1.325 | 0.007 | * | 0.004 | * |
| Ln(VIEW_LAKE) | 0.016 | 15.947 | *** | 1.489 | 0.016 | *** | 0.012 | *** |
| Ln(POP_DENS) | -0.016 | -9.103 | *** | 1.110 | -0.018 | *** | -0.014 | *** |
| Ln(CATERING) | 0.021 | 11.113 | *** | 2.385 | 0.021 | *** | 0.014 | *** |
| FOREIGNER | -0.002 | -3.921 | *** | 1.181 | -0.002 | ** | -0.002 | *** |
| GROCERY | 0.009 | 2.230 | ** | 1.212 | 0.008 | *** | 0.005 | |
| CONSTRUCT | 0.001 | 8.718 | *** | 1.961 | 0.001 | *** | 0.001 | *** |
| Ln(INCOME) | 0.236 | 17.919 | *** | 1.511 | 0.234 | *** | 0.163 | *** |
| λ | | | | | 0.318 | *** | | |
| ρ | | | | | | | 0.301 | *** |
| R square[1] | 0.511 | | | | 0.536 | | 0.538 | |
| Log-Likelihood | | | | | 3742 | | 3725 | |
| Breusch-Pagan | 204.4 | | *** | | 326.7 | *** | 332.9 | *** |

$N$ =9218; Probability of rejecting $H_0$ = *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

[1] for OLS it is adjusted $R^2$, for SEM/SAR it is pseudo-$R^2$

# 5. Conclusions

In this paper, a hedonic rent price model has been estimated based on publicly available real estate data. The data has been matched with a broad set of spatial variables, including those generated by advanced GIS techniques. A considerable wide set of variables proved to be significant in an OLS model while controlling for multicorrelinearity. Even with various spatial independent variables included, tests for spatial autocorrelation indicated spatial dependence of the rent price. Consequently, after selecting a reasonable contiguity matrix, both a SEM and a SAR model have been estimated, which achieve a higher $R^2$ as the OLS model. Despite of the improvement of the model, heteroscedasticity remains an issue and will be solved in the future through methodological improvement.

Moreover, there is recent work, which proposes to advance the shown approach of SEM and SAR models further (Anselin, 2003). The test of appropriateness of those models for the dataset under study remains to be done.

# 6. References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1992) *SpaceStat Tutorial – A Workbook for Using SpaceStat in the Analysis of Spatial Data*, University of Illinois, Urbana.

Anselin, L. (2002) Under the hood – Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, **27** (3) 247-267.

Anselin, L. (2003) Spatial externalities, spatial multipliers, and spatial econometrics, *International Regional Science Review*, **26** (2) 152-166.

Anselin, L. (2005) *Exploring spatial data with GeoDA: a workbook*, Spatial Analysis Laboratory, Department of Geography, University of Illinios, Urbana-Champaign.

Anselin, L. (2006a) Spatial econometrics, in T. Mills and K. Patterson (eds.) *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*, 901-969, Palgrave Macmillan, Basingstooke.

Anselin, L. (2006b) Spatial Regression, http://www.sal.uiuc.edu/stuff/stuff-sum/pdf/anselin_spareg.pdf, August 2007.

Baranzini, A. and C. Schaerer (2007) *A sight for sore eyes – Assessing the value of view and landscape use on the housing market*, HEG, Geneva.

Basu, S. and T.G. Thibodeau (1998) Analysis of spatial autocorrelation in house prices, *Journal of Real Estate Finance and Economics*, **17** (1) 61-85.

Bivand, R. (2002) Spatial econometrics functions in R: Classes and methods, *Journal of Geographical Systems*, **4** (4) 405-421.

Bivand, R.S. and B.A. Portnov (2004) Exploring spatial data analysis techniques using R: The case of observations with no neighbors, in: L. Anselin, R.J.G.M. Florax and S.J. Rey (eds.) *Advances in Spatial Econometrics*, 121-142, Springer Verlag, Berlin.

Bourassa, S.C., M. Hoesli and J. Sun (2003) What's in a view?, International Center for Financial Asset Management and Engineering, *Research Paper*, **79**, University of Geneva, Geneva.

Bourassa, S.C., E. Cantoni and M. Hoesli (2005) Spatial dependence, housing submarkets, and house prices, International Center for Financial Asset Management and Engineering, *Research Paper*, **151**, University of Geneva, Geneva.

Bozdogan, H. (1987) Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions, *Psychometrika,* **52** (3) 345-370.

Can, A. (1990) The measurement of neighborhood dynamics in urban house prices, *Economic Geography*, **66** (3) 254-272.

Dubin, R., R.K. Pace and T.G. Thibodeau (1999) Spatial autoregression techniques for real estate data, *Journal of Real Estate Literature*, **7** (1) 79-95.

Fletcher, M., P. Gallimore and J. Mangan (2000) Heteroscedasticity in hedonic house price models, *Journal of Property Research*, **17** (2) 93-108.

Griffith, D. A. (2003) Spatial *autocorrelation and spatial filtering*, Springer-Verlang, Berlin.

Hackney, J.K., M. Bernard, S. Bindra and K.W. Axhausen (2006) Explaining road speed with spatial lag and spatial error regression models, Paper presented at the *86$^{th}$ Annual Transportation Research Board Meeting*, Washington D.C.

Ihaka, R. and R. Gentleman (1996) R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, **5** (3) 299-314.

Jud, G.D., J.D. Benjamin and G.S. Sirmans (1996) What do we know about apartments and their markets?, *The Journal of Real Estate Research*, **11** (3) 243-257.

Kim, C.W., T.T. Phipps and L. Anselin (2003) Measuring the benefits of air quality improvement: a spatial hedonic approach, *Journal of Environmental Economics and Management*, **45** (1) 24-39.

Lancaster, K. J. (1966) A new approach to consumer theory, *Journal of Political Economy*, **74**, 132-157.

Lancaster, K.J. (1971) *Consumer Demand: A New Approach*, Columbia University Press, New York.

Löchl, M., M. Bürgle and U. Waldner (2007) Handbuch Simulationsprojekt Grossraum Zürich, Arbeitsberichte *Polyprojekt Zukunft urbane Kulturlandschaften*, **10**, NSL, ETH Zürich, Zürich.

Löchl, M. (2006) Real estate and land price models for UrbanSim’s Greater Zurich application, Arbeitsberichte *Polyprojekt Zukunft urbane Kulturlandschaften*, **6**, NSL, ETH Zürich, Zürich.

Maddala, G.S. (2001) *Introduction to Econometrics*, Third Edition, Wiley, West Sussex.

Marks, D. (1984) The effect of rent control on the price of rental housing: an hedonic approach, *Land Economics*, **62** (1) 81-94.

Pace, R.K., R. Barry and C.F. Sirmans (1998) Spatial statistics and real estate, *Journal of Real Estate and Economics*, **17** (1) 5-13.

Rieder, T. (2005) *Die Bestimmung der Mieten in der Schweiz*, Lizenzariatsarbeit, Volkswirtschaftliches Institut, Abteilung Ökonometrie, Universität Bern.

Rosen, S. (1974) Hedonic prices and implicit markets: product differentiation in price competition, *Journal of Political Economy*, **82** (1) 35-55.

Salvi, M. and P. Schellenbauer (2004) *Preise, Mieten und Renditen: Der Immobilienmarkt transparent gemacht*, Zürcher Kantonalbank, Zürich.

Savin, N.E. and K.J. White (1977) The Durbin-Watson test for serial correlation with extreme sample size or many regressors, *Econometrica*, **45** (8) 1989-1996.

Scognamiglio, D.F. (1999) *Methoden zur Immobilienbewertung im Vergleich – Eine empirische Untersuchung für Schweizer Immobilien*, Dissertation, Universität Bern, Bern.